A3273-8

MAY 1990

ACADEMIC MEDICINE

Recruiting Black Medical Students

Assessment

- Use of Standardized Patients
- Measuring Program Outcomes
- Clinical Skills at Four Schools
- Medicine Clerks

Implications of Primary Health Care

Medical Informatics at Dartmouth

Problems of Late Hospitalization

TIME Patient Simulation Model

Integrating Clinical Data

Family Medicine Training

VOLUME 65 • NUMBER FIVE

MAY 1990
VOLUME 65
NUMBER FIVE

ACADEMIC MEDICINE

JOURNAL OF THE
ASSOCIATION OF
AMERICAN MEDICAL
COLLEGES

Invited Articles	Recruiting Black Medical Students: A Decade of Effort, *by Robert E. Taylor, Jr., Ph.D., James C. Hunt, M.D., and Patra B. Temple, Ed.D.*	279
	Results of a Survey on the Use of Standardized Patients to Teach and Evaluate Clinical Skills, *by Paula L. Stillman, M.D., Mary Beth Regan, Ed.D., Mary Philbin, and Heather-Lyn Haley*	288
	The Measurement of Outcomes in the Assessment of Educational Program Effectiveness, *by Donald G. Kassebaum, M.D.*	293
Ideas for Medical Education	Using Classic and Contemporary Visual Images in Clinical Teaching, *by Janine C. Edwards, Ph.D.*	297
Medical Informatics	Medical Informatics for the Other Ninety Percent: The Dartmouth Experience, *by J. Robert Beck, M.D., Edward K. Shultz, M.D., M.S., and Brian R. Edwards, Ph.D., MS*	298
International Medical Education	Primary Health Care: Implications for the Medical Profession and Education, *by Snehendu B. Kar, Dr.P.H., M. Sc.*	301
Editorial	On Standardized Patients and Clinical Skills Assessment, *by Donald G. Kassebaum, M.D.*	307
Letters to the Editor	Residents' Hours and Supervision, *by Michael D. Cusimano, M.D., and William S. Tucker, M.D.*	308
	Specialty Choice, *by Janet Bickel, M.A.*	308
	Predicting Matriculation, *by Marc S. Nelson, M.D.*	308
From the President		310
AAMC Focus		311

National Policy Perspectives 312

Research Reports
ARTICLES

The Problem of Late Hospitalization: A Quality and Cost Issue, *by Joseph S. Gonnella, M.D., Daniel Z. Louis, M.S., Carter Zeleznik, Ph.D., and Barbara J. Turner, M.D.* 314

An Assessment of the Clinical Skills of Fourth-Year Students at Four New England Medical Schools, *by Paula L. Stillman, M.D., Mary Beth Regan, Ed.D., David B. Swanson, Ph.D., Susan Case, Ph.D., John McCahan, M.D., Joel Feinblatt, Ph.D., Stephen R. Smith, M.D., Janice W. Willms, M.D., and D. Virginia Nelson* 320

A Field Test of the TIME Patient Simulation Model, *by William G. Harless, Ph.D., Robert C. Duncan, Ph.D., Marcia A. Zier, William R. Ayers, M.D., Jerome R. Berman, M.D., and Henry S. Pohl, M.D.* 327

A Comparison of Two Methods for Facilitating Clinical Data Integration by Medical Students, *by I. Jon Russell, M.D., Ph.D., William D. Hendricson, M.S., Gary D. Harris, M.D., and Denise V. Gobert* 333

Evaluation of Students in Medicine Clerkships, *by Gregory J. Magarian, M.D., and Dennis J. Mazur, M.D., Ph.D.* 341

A New Faculty Training Model for Countries Developing Academic Family Medicine Programs, *by Douglas H. Johnson, M.D., and Alfred Zammit-Montebello, M.Ed.* 346

SHORT PAPER

The Availability of Fellowship Training for Foreign Family Physicians, *by Michael D. Cohen, M.B., Ch.B., and Joel H. Merenstein, M.D.* 348

Publishing Services: Hanley & Belfus, Inc.
210 South 13th Street, Philadelphia, PA 19107

MAY 1990
VOLUME 65
NUMBER FIVE

ACADEMIC MEDICINE

JOURNAL OF THE
ASSOCIATION OF
AMERICAN MEDICAL
COLLEGES

EDITOR: Addeane S. Caelleigh

DEPUTY EDITOR: Albert Bradford

SUBMISSIONS EDITOR: Christopher Greene

EDITORIAL ASSISTANT: Lisa R. Dittrich

PUBLISHER'S EDITOR: Sandra A. Lovegrove

ASSOCIATE EDITORS
Senior Associate Editor
Donald G. Kassebaum, M.D.
Edward J. Stemmler, M.D.
Association of American Medical Colleges

Ideas in Medical Education
Janine C. Edwards, Ph.D.
St. Louis University, School of Medicine

International Medical Education
David S. Greer, M.D.
Brown University Program in Medicine

Informatics
Daniel R. Masys, Ph.D.
Lister Hill National Center for Biomedical Communications

Book Reviews
Sherman Mellinkoff, M.D.
University of California, Los Angeles School of Medicine

Academic Computing in Medicine
Robert F. Rubeck, Ph.D.
Arizona Health Sciences Center

EDITORIAL BOARD
Chairman
Milton Corn, M.D.
Georgetown University Hospital
Philip Anderson, M.D.
University of Missouri-Columbia School of Medicine
Nancy Bennett, Ph.D.
Harvard Medical School
Eta Berner, Ed.D.
University of Alabama School of Medicine
Ruth Bulger, Ph.D.
National Academy of Sciences
Vincent A. Fulginiti, M.D.
Tulane University School of Medicine
Nancy Gary, M.D.
Albany Medical College
Beatrix Hamburg, M.D.
Mt. Sinai Hospital
Murray M. Kappelman, M.D.
The University of Maryland School of Medicine
Ralph W. Muller
University of Chicago Hospitals
Herbert Pardes, M.D.
College of Physicians and Surgeons of Columbia University
James A. Pittman, Jr., M.D.
University of Alabama School of Medicine
W. Vickery Stoughton
Toronto General Hospital
Martha Vaughan, M.D.
National Institutes of Health

OFFICERS OF THE ASSOCIATION
Chairman
David H. Cohen, Ph.D.
Northwestern University

Chairman-Elect
William T. Butler, M.D.
Baylor College of Medicine

Immediate Past Chairman
D. Kay Clawson, M.D.
University of Kansas Medical Center School of Medicine

President
Robert G. Petersdorf, M.D.

Chairman, Council of Deans
L. Thompson Bowles, M.D., Ph.D.
George Washington University School of Medicine and Health Sciences

Chairman, Council of Academic Societies
Joe Dan Coulter, Ph.D.
University of Iowa College of Medicine

Chairman, Council of Teaching Hospitals
Raymond G. Schultze, M.D.
University of California, Los Angeles Medical Center

Academic Medicine (ISSN 1040-2446) is owned and published monthly by the Association of American Medical Colleges, One Dupont Circle, NW, Washington, DC 20036. Publishing services by Hanley & Belfus, Inc., 210 South 13th Street, Philadelphia, Pennsylvania 19107. Second-class postage paid at Washington, DC and additional mailing offices.

Copyright 1990 by the Association of American Medical Colleges. All material subject to this copyright may be photocopied for the noncommercial purpose of scientific or educational advancement. Printed in the U.S.A.

Subscription rates: U.S., Canada, Latin America—$60.00 per year and $7.00 per single copy ($30.00 per year for students); all other locations—$70.00 per year and $8.00 per single copy.

All advertising material published in this journal is intended to conform to ethical standards, but acceptance does not imply endorsement. The views expressed in all articles and features in this journal are those of the authors and do not necessarily reflect the position of the AAMC.

CORRESPONDENCE
Manuscripts: The Editor, *Academic Medicine*, Association of American Medical Colleges, One Dupont Circle, NW, Washington, DC 20036.

Advertising: Diane Sherel, Advertising Manager, Hanley & Belfus, Inc., 210 South 13th Street, Philadelphia, PA 19107; (215) 546-0313.
Subscriptions and changes of address: Membership and Publication Orders, AAMC, One Dupont Circle, NW, Washington, DC 20036; (202) 828-0416.
Cancellations of an *Academic Medicine* subscription must be submitted in writing to Publication Orders, AAMC, One Dupont Circle, NW, Washington, DC 20036. Claims for back issues will be honored for only the three months before the cancellation.
Reprints: Request reprints directly from the authors of articles, not from *Academic Medicine*.
POSTMASTER: Send address changes to *Academic Medicine*, AAMC, Suite 200, One Dupont Circle, NW, Washington, DC 20036.

5. *Minority Students in Medical Education: Facts and Figures IV.* Washington, D.C.: Association of American Medical Colleges, 1988.
6. *Statistical Abstract of the United States: 1987.* Washington, D.C.: U.S. Bureau of the Census, 1987.
7. Personal communication, June 13, 1988. Tennessee Department of Vital Statistics, Nashville, Tennessee.
8. Personal communication, September 6, 1989. Tennessee Department of Health and Environment, Division of Information Resources, Nashville, Tennessee.
9. Personal communication, June 21, 1988. Association of American Medical Colleges, Student and Applicant Information Management System Data Base.
10. Etzel, S. I. Medical Education in the United States 1978-1979. *JAMA* **243**(1980):841-866.
11. *Desegregation Monitoring Report.* Nashville, Tennessee: Tennessee Higher Education Commission, April 1989.
12. Personal communication, June 15, 1988. Association of American Medical Colleges, Faculty Roster System Data Base.
13. Minority Admission Score Card. *The New Physician.* **35**(December 1986): 20-21.

PAULA L. STILLMAN, M.D., MARY BETH REGAN, Ed.D., MARY PHILBIN, and HEATHER-LYN HALEY

Results of a Survey on the Use of Standardized Patients to Teach and Evaluate Clinical Skills

Abstract—In 1989, a survey was sent to each U.S. and Canadian medical school requesting information about how standardized patients are used for teaching and evaluating clinical skills, and 95% of the schools responded. Although there was widespread use of standardized patients throughout the curricula, the role and training of these patients varied markedly within a given school as well as across schools. One outcome of this survey is the development of a network to share resources, protocols, and training material to enhance the development of this educational strategy. *Acad. Med.* 65(1990):288-292.

Standardized patients (SPs) have been used throughout medical school curricula for at least 25 years[1-8] to teach and evaluate clinical skills. SPs, often called simulated patients, patient instructors, or programmed patients, are lay persons trained to replicate a clinical encounter consistently and realistically. They may have no specific medical background, may be actors, or may be actual patients with or without real physical abnormalities. The SP may only simulate a patient or, with additional training, may record aspects of the student's performance and evaluate the student's skills on specially designed forms. With further training, the SP may function as a teacher and at the end of the encounter provide the student with individualized and constructive feedback. According to the results of the 1989 LCME Part II Medical Education Questionnaire, SPs are being used at an increasing number of schools.[9] Although a majority of U.S. medical schools indicate that they use standardized patients in the curriculum, the extent to which and purposes for which they are used are not generally known. Many medical educators believe that SPs are a valuable tool in a variety of clinical exercises to provide students with uniform teaching and evaluation of basic clinical skills and to ensure consistent exposure to common ambulatory problems. One of the purposes of this survey was to document the extent of this use and to assess the interest of medical schools in establishing a network that would facilitate the development or expansion of SP programs.

Dr. Stillman is curriculum dean, Dr. Regan is an education specialist, Ms. Philbin is coordinator of the standardized patient program, and Ms. Haley is a research assistant, all at the University of Massachusetts Medical Center, Worcester.

Correspondence and requests for reprints should be addressed to Dr. Stillman, Office of the Curriculum Dean, University of Massachusetts Medical Center, 55 Lake Avenue North, Worcester, MA 01655.

Related articles about standardized patients appear on pages 307, 320, and 327.

Method

In July 1989, a survey on the use of standardized patients for teaching and evaluating medical students and residents was mailed to 142 curriculum deans at U.S. and Canadian medical schools. After five weeks, a second copy of the survey was mailed to all that had not replied. A total of 136 (95%) responded to the survey. (The detailed results are available from the first author.) The survey questionnaire requested detailed documentation on the use of SPs throughout the curriculum. Data were collected to answer these questions: (1) How are SPs used at various institutions to teach and evaluate through the curriculum? (2) What are common problem areas? (3) What are some of the characteristics of successful programs? (4) What cases, evaluation instruments, etc., are there that can be

shared? and (5) Is there interest enough for a resource network to be established among academic institutions?

Results

The results are summarized here in terms of 16 topics. Although the survey results are from 1989, the findings can be considered current because SP programs tend to be stable and long-term.

Use of SPs. Standardized patients are used in some way at 94 (70%) of the 136 schools that responded to the survey. The survey results presented in Figure 1 illustrate the numbers of schools using SPs across the curriculum for teaching and evaluation. The numbers on the graphs are higher than the total number of schools listed in the text because SPs are often used in more than one curriculum year.

Breast and Pelvic Examination. The most popular use of SPs is to train students in breast and pelvic examination. This type of program now exists at 84 medical schools. It mainly occurs in the second and third years of the curriculum, with SPs used primarily to teach. Each teaching session lasts 10 to 62 minutes. The schools that use SPs to evaluate students or residents allot 5 to 25 minutes for that portion of the encounter. Figure 1A summarizes these uses of SPs.

Male Genitourinary (GU) Examination. SPs are used for the male GU exam at 45 schools, mostly in the second year and primarily for teaching purposes. The length of encounter is between 10 and 40 minutes, and evaluation of the encounter usually 5 to 17 minutes. Figure 1B details the use of SPs for this examination.

Interviewing Skills. Sixty-two schools use SPs for teaching or evaluating interviewing skills. The category "interviewing skills" was defined on the questionnaire as the process by which information is obtained from or shared with the patient. The length of the SP-student encounter is between 17 and 29 minutes, with an additional 6 to 25 minutes being spent on evaluation. A breakdown of SP use by year of training, as well as whether the SP is used for teaching or evaluating the interviewing process, can be seen in Figure 1C.

History Content. Fifty-seven schools use their standardized patients to help the students learn to take a relevant medical history. History content refers to the actual medical history data obtained and does not take into consideration interviewing or interpersonal skills, although overlap certainly exists. The most popular history content areas are the adult, geriatric, and psychiatric histories. The SP-student encounter varies between 14 and 37 minutes, with an additional 7 to 19 minutes for evaluation. The use of SPs for teaching and evaluating history content is illustrated in Figure 1D.

Patient Education and Counselling. Thirty-nine schools use SPs to teach or evaluate counselling skills and patient education. This is defined as the delivery to the patient of information such as the need for behavior modification or lifestyle changes, counselling, or delivering bad news to a patient or the family. The length of the encounter varies from 17 to 32 minutes, with an evaluation time between 10 and 25 minutes. The pattern of usage, with the primary focus on second-year students and residents, can be seen in Figure 1E.

Complete Physical Examination. SPs are used to teach or evaluate the complete physical examination at 33 of the responding schools. The encounter lasts between 20 and 56 minutes, with an additional 5 to 44 minutes spent on evaluation. SPs are used most often in the second year of the curriculum for both teaching and evaluating the complete physical examination, as can be seen in Figure 1F. This aspect of the program is also used for evaluating residents' physical examination skills.

Segments of Physical Examination. Forty-eight schools reported using SPs for teaching and/or evaluating segments of the physical examination. The encounter between the student and the standardized patient lasts between 11 and 30 minutes, with 5 to 26 minutes set aside for evaluation. Figure 1G details the use of SPs in this area, showing that they are used primarily for second- and third-year students.

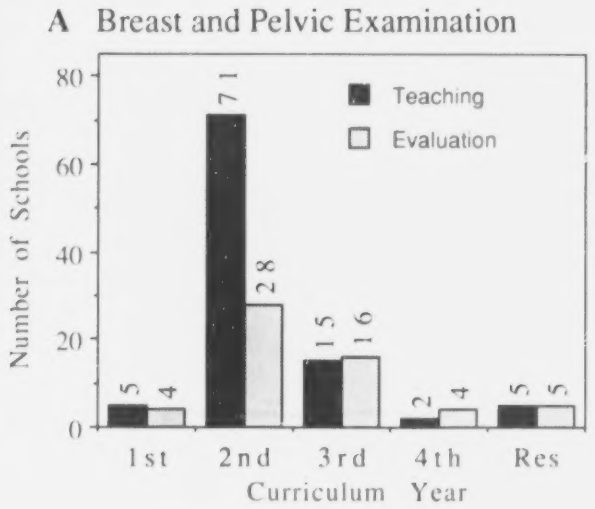
Focused Encounter. Forty-six of the schools surveyed use SPs for focused encounters (brief histories and physical examinations). They are most frequently used for both teaching and evaluation of second-year students. The emphasis shifts to evaluation for third- and fourth-year students and residents, as shown in Figure 1H. The length of encounter is between 14 and 38 minutes, with 9 to 22 minutes spent on evaluation.

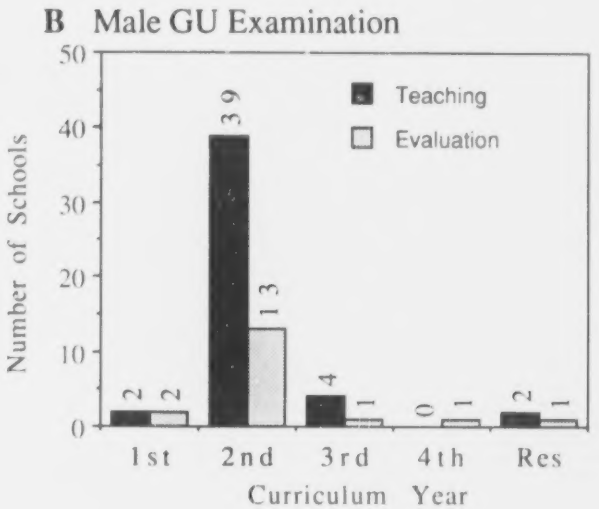
Checklist. A majority of schools (64) that use standardized patients have checklists to document students' performance. Depending on the type of encounter, the checklist may be completed by a faculty member, the SP, or both (see Table 1).

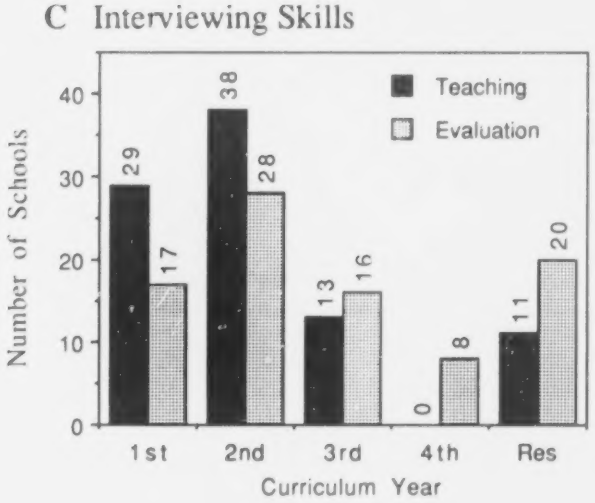
Feedback. Feedback is given to students who interact with standardized patients at 87 of the 94 schools that have such a program in place. Ninety-eight percent of the schools provide oral feedback, while 38% give written feedback; 93% give the student feedback immediately following the encounter. Depending on the type of encounter, the feedback may be given by a faculty member, the SP, or both. This information is also summarized in Table 1.

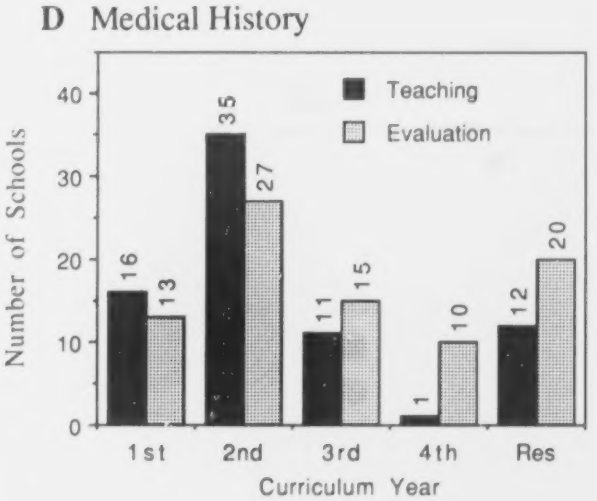
Videotape Use. Forty-three schools videotape encounters with standardized patients. At 33 of these schools, the tapes are used for evaluation of individual students, while 23 schools use them for group teaching and 18 schools, to train new SPs. A majority of taping is done with first-year students and residents. The most popular areas for videotaping are the history content, patient education and counselling, and focused history and physical examinations.

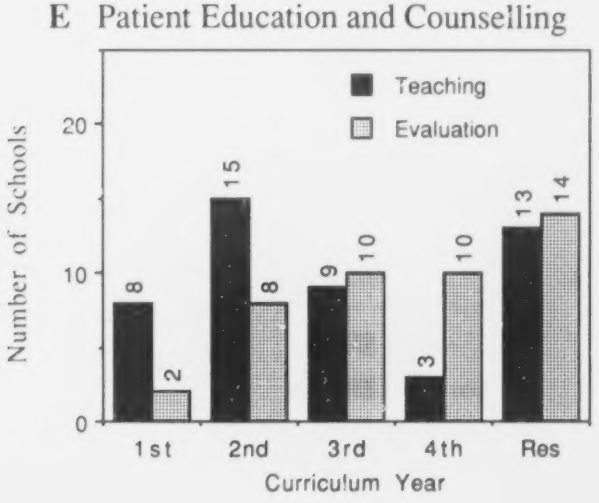
Training of Standardized Patients. Standardized patients are trained by a non-M.D. trainer at 68 of the 94 schools that use SPs, 53 use M.D. trainers, and 27 use the case author. Again, the totals are higher than the number of schools because more than one type of trainer may be used at a single school. The most common training technique is some combination of the individual session (69

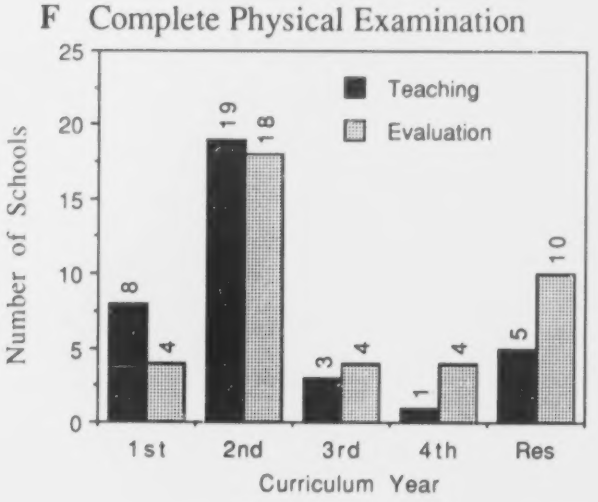

A Breast and Pelvic Examination
Number of Schools
Teaching
Evaluation
1st 2nd 3rd 4th Res
Curriculum Year


B Male GU Examination
Number of Schools
Teaching
Evaluation
1st 2nd 3rd 4th Res
Curriculum Year


C Interviewing Skills
Number of Schools
Teaching
Evaluation
1st 2nd 3rd 4th Res
Curriculum Year


D Medical History
Number of Schools
Teaching
Evaluation
1st 2nd 3rd 4th Res
Curriculum Year


E Patient Education and Counselling
Number of Schools
Teaching
Evaluation
1st 2nd 3rd 4th Res
Curriculum Year


F Complete Physical Examination
Number of Schools
Teaching
Evaluation
1st 2nd 3rd 4th Res
Curriculum Year

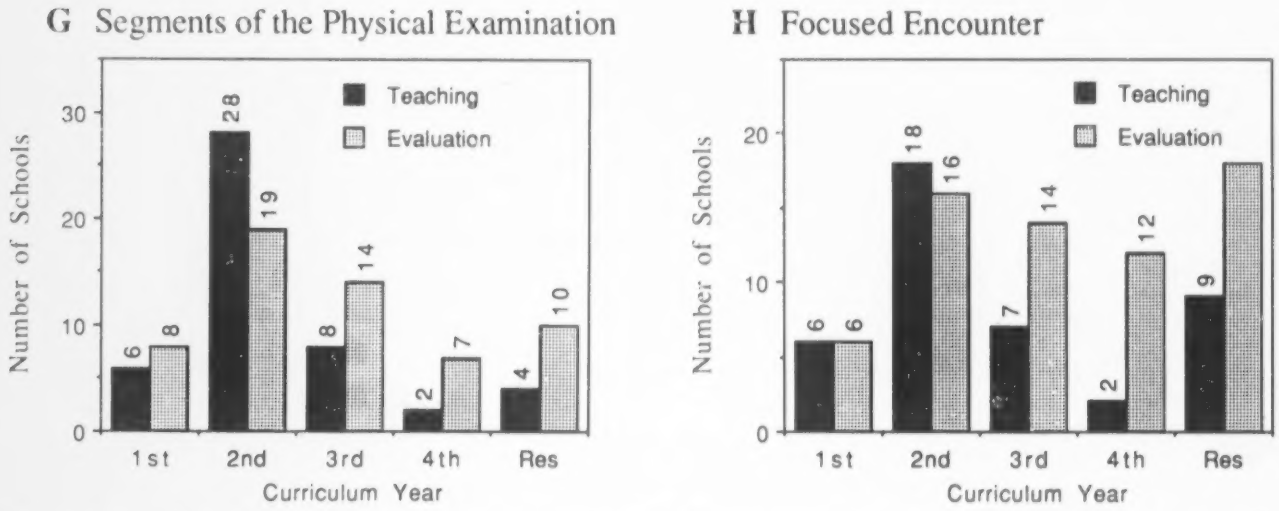


Figure 1 (facing page and above). **Numbers of schools among 136 surveyed in 1989 that used standardized patients for teaching and evaluation in various programs.**

schools) and group session (64 schools), while 41 schools rely on videotapes. Direct observation of other SPs is commonly employed by 42 schools, while 41 schools teach from manuals; only six schools reported training SPs by having them directly observe actual patients.

The average program spends seven hours training the SP to provide a medical history and simulate symptoms or an affective disorder, four hours to teach an SP a focused physical examination, eight hours for the breast and pelvic examination, and six hours for the male GU examination. It takes approximately ten hours to train an SP to evaluate the student's interviewing skills and to provide adequate feedback and teaching. Fifteen hours are required to train SPs to evaluate the student's physical examination skills, including teaching the SPs to provide feedback and teaching in this area.

Administration of Program, Faculty Participation, and Recruitment. Thirty-nine schools with SP programs have a central office that coordinates the program, while 55 do not. The numbers of SPs enrolled in individual programs vary from 2 to 125, with a mean of 27, and the numbers of case protocols range from zero to 400, with a mean of 39. Although one school has used SPs for 25 years, they have been used for a mean of four years at most schools.

Faculty participation is high, with 70 schools reporting that their faculty observe encounters; at 63 schools faculty provide feedback. At 61 schools faculty train the standardized patients, and at 53 schools faculty write cases. Recruitment methods are similar overall, with 53 schools using word of mouth as their main method. Other sources for finding SPs include personal contact, medical and nursing students, senior citizens' groups, campus newspapers and bulletin boards, community support groups such as Alcoholics Anonymous, and drama and acting groups.

Strengths and Weaknesses. Some of the major strengths of the programs surveyed were strong student and faculty support, the importance of "real" patient feedback for students, the opportunity for students to practice counselling skills, and strong case development. Weaknesses mentioned were programs with a limited scope and wasted potential, the need for better defined, stronger evaluation strategies, a lack of organization and time, and inadequate financial support.

Start-up Concerns. For those schools that do not have an SP program, the largest start-up concern is financial, with 28 schools citing it as an obstacle. Fifteen have a problem

Table 1

Numbers of U.S. Medical Schools Using Checklists and Giving Feedback to Students during the Use of Standardized Patients to Teach Eight Skills, 1989

Skills	Total No.	Checklists Used			Feedback Given		
		No.	By Faculty	By SP	No.	By Faculty	By SP
Interviewing skills	62	51	34	37	55	48	39
Medical history content	57	48	31	34	48	43	29
Patient education	39	31	14	23	31	22	18
Complete physical examination	33	27	17	15	23	15	17
Segments of physical examination	48	34	23	17	34	28	18
Breast and pelvic examination	84	35	13	27	68	35	58
Male genitourinary examination	45	19	8	13	34	17	25
Focused encounter	46	34	24	21	31	24	15

with lack of interest from the faculty, while only six schools list administrative lack of interest as a drawback. Fifteen schools lack individuals who have the skills to run the program and train SPs, and nine cite an inability to locate appropriate standardized patients. Another concern is an overall lack of faculty time to devote to developing and sustaining such a program.

Formation of Network. Ninety-one of the 94 schools with an SP program are interested in forming a network. Forty-four schools have cases to share, while 14 offered training videotapes and 15 indicated that they had training manuals. Other materials respondents are willing to share include checklists, objectives, lecture materials, facilities, evaluation forms, and SP handbooks.

Discussion

It is clear from reading the survey results that the term standardized patient implies a variety of entities. Because there is no uniformity in defining the role or training of the SP, it is important that we develop a common classification system so that communication within and across institutions can be enhanced. The term is now used to include the spectrum of real patients without training, simulated patients, and role playing; the function of SPs ranges from simulator, to evaluator, to teacher. There is a pressing need for a hierarchical classification with agreement as to the meanings of specific terms.

We are enthusiastic about the frequent use of SPs to supplement current teaching activities and as part of a diagnostic assessment of students' skills throughout the medical school curricula. Because of the changing nature of the hospitalized patient population and the difficulty in teaching in an ambulatory setting, SPs can be used to ensure that each student has exposure to common ambulatory problems, such as those of patients with newly diagnosed diabetes or anginal chest pain. They can help ensure uniformity in curriculum. They can also be used to teach basic clinical skills such as interviewing and physical examination techniques, and clinicians' time can then be reserved for discussing diagnostic clinical problem solving once basic skills are mastered. SPs are also an excellent addition to problem-based small-group teaching. The cases discussed can come alive by having students interact with an SP who portrays the patient realistically. Several SPs can be trained in the same role, allowing each small group the benefit of this exposure.

Some types of SP cases seem more appropriate for teaching, while others seem better suited for evaluation. For example, asking an SP for permission to perform an autopsy on the next of kin or informing an SP that a malignancy has recurred provides the student an opportunity to practice these very difficult communication skills in a safe setting. Common ambulatory problems are more appropriate for evaluating clinical skills such as data gathering. A very powerful use of this technique is to allow a student to interact with an SP and then to have the SP evaluate the student's performance and provide immediate feedback, discussing strengths and weaknesses in the student's performance. The SP can then return to his role, and the student can practice what he has just learned.

Many schools appear strongly interested in expanding their SP programs. A major barrier is the lack of trained staff to develop protocols and train SPs. The development of a network for sharing resources and conducting workshops on case development and training of SPs is now under way and should decrease this problem. We do not believe that SPs can ever replace the rich encounter that occurs between a student, a faculty member, and an actual patient, but they can be a valuable adjunct to teaching.

References

1. Stillman, P. L., Ruggill, J. S., and Sabers, D. L. The Use of Practical Instructors to Teach and Evaluate a Complete Physical Examination. *Eval. Health Prof.* **1**(1978): 49–54.
2. Stillman, P. L., Burpeau-DiGregorio, M. Y., Nicholson, G. I., Sabers, D. L., and Stillman, A. E. Six Years of Experience Using Patient Instructors to Teach Interviewing Skills. *J. Med. Educ.* **58**(1983)941–946.
3. Stillman, P. L., et al. Assessing Clinical Skills of Residents with Standardized Patients. *Ann. Intern. Med.* **105**(1986): 762–771.
4. Stillman, P. L., and Swanson, D. B. Insuring the Clinical Competence of Medical School Graduates Through Standardized Patients. *Arch. Intern. Med.* **147**(1987): 1049–1052.
5. Stillman, P. L., Regan, M. B., and Swanson, D. B. A Diagnostic Fourth-Year Performance Assessment. *Arch. Intern. Med.* **147**(1987):1981–1985.
6. Stillman, P. L., et al. An Assessment of the Clinical Skills of Fourth-Year Students at Four New England Medical Schools. *Acad. Med.* **65**(1990):320–326.
7. Kretzschmar, R. M. Evolution of the Gynecology Teaching Associate: An Education Specialist. Am. J. Obstet. Gynecol. **131** (1978):367–375.
8. Barrows, H. S. *Simulated (Standardized) Patients and Other Human Simulations.* Chapel Hill, North Carolina: Health Sciences Consortium, 1988.
9. Personal communication, Donald G. Kassebaum, M.D., director of accreditation, Association of American Medical Colleges, Washington, D.C., December 1989.

The Problem of Late Hospitalization: A Quality and Cost Issue

JOSEPH S. GONNELLA, M.D., DANIEL Z. LOUIS, M.S., CARTER ZELEZNIK, Ph.D., and BARBARA J. TURNER, M.D.

Abstract—Emphasis on controlling health-care costs has led to many activities aimed at avoiding medically unnecessary hospitalizations. Much less attention has been given to patients hospitalized in advanced stages of illness and the impact of these late admissions on cost and quality of care. A panel of physicians developed criteria to categorize hospital admissions into one of three groups—early, timely, or late—based on the timing of the initial hospitalizations of patients admitted with any one of 14 diagnoses. Over a period of one year (fiscal year 1984) the criteria were applied retrospectively to 2,713 patients admitted to either of two hospitals. Twenty-one percent of the admissions studied in one hospital and 19% in the other were judged to occur later than was desirable. The mean length of stay for late hospitalizations exceeded that for timely hospitalizations by 11.1 days at one hospital and by 7.5 days at the other ($p < .01$). Similar patterns were observed in analyzing the 14 diseases individually and in an analysis of hospital charges at the one hospital where charge data were available. In-hospital mortality rates for patients with a principal diagnosis of bacterial pneumonia were over ten times greater for those admitted late than for those whose admissions were timely (39.0% versus 3.8%, $p < .001$, at one hospital; 28.9% versus 2.1%, $p < .001$, at the other). While not all late hospitalizations are avoidable, the authors believe that the analysis of late hospitalization patterns is an important part of any effort that can be made to reduce them. *Acad. Med.* 65(1990):314–319.

Current utilization review activities focus on medically unnecessary hospitalizations or those considered longer than necessary.[1–3] With this focus, these programs ignore or underestimate the problem of patients who are admitted to the hospital in advanced stages of diseases which potentially could have been detected and treated earlier. Changes in patient care reimbursement methods and in the structure of health care delivery organizations make it even more urgent to monitor hospitalizations of patients in late stages of disease.[4–6]

This study was designed to determine whether significant opportunities exist to reduce patient morbidity and health care costs by utilizing health care services before diseases progress to advanced stages.

Operationally, this study is based on the "disease staging" concept.[7,8] Staging is a disease classification system that focuses specifically on the biological dimension of severity of illness (etiology, pathophysiology, and ranking of local and systemic complications). Over the past ten years, staging criteria have been developed for 400 disease entities.[9] Disease staging previously has been used to evaluate the comparative costs of caring for patients with early and more advanced stages of disease. Analysis of over 30,000 Medicare discharges has shown that hospitalizations of advanced-stage diabetic patients cost 50% more than do those for patients in early stages.[10]

Method

In order to modify disease staging for the current project, a panel of physicians (selected from a number of departments at one medical school) divided the staging criteria for each of 14 diseases into three groups in terms of the timing of hospitalization:

1. *Early.* Accepted level of medical care can be provided in an outpatient setting. The admission is premature or inappropriate. (*Example*: admission of a patient with chronic cholecystitis who is treated only with oral medication and undergoes no surgical procedure.)
2. *Timely.* The patient is admitted at the earliest detectable stage at which the probability of disease progression can be significantly reduced through use of resources available only in a hospital setting; or the patient has another condition(s) unrelated to the principal diagnosis that makes ambulatory treatment inappropriate; or patient compliance is such that inpatient care is necessary. (*Example*: admission of a patient with acute cholecystitis without perforation of the gallbladder.)

Dr. Gonnella is senior vice president and dean, Jefferson Medical College, Philadelphia, Pennsylvania. Mr. Louis is managing director, Dr. Zeleznik is a senior researcher, and Dr. Turner is a research associate; all at the Center for Research in Medical Education and Health Care, Jefferson Medical College.

Correspondence and requests for reprints should be addressed to Mr. Louis, Jefferson Medical College, 1025 Walnut Street, Room 119, Philadelphia, PA 19107.

3. *Late.* Hospitalization is delayed. The patient has complications that could have been prevented with earlier hospitalization. (*Example*: admission of a patient with acute cholecystitis who has gangrene of the gallbladder.)

It is important to stress that not all late hospital admissions are preventable. Late hospitalization can result from patient-related problems (such as failure to seek care or failure to follow a prescribed treatment); provider problems (such as an incorrect or late diagnosis or treatment); or system problems (such as lack of access to care or an overzealous admission certification program). Also, in some cases, a disease may progress so rapidly that nothing could reasonably have been done to avoid the disease progression prior to hospitalization.

In this study the timing of hospitalization for 14 diseases was examined in two hospitals, a university-based hospital in a major metropolitan area (Hospital A) and a community-based hospital in a small city (Hospital B). Data were collected on all initial hospitalizations in fiscal year 1984 for 14 high-frequency diagnoses. These conditions were chosen to assure analysis of diseases involving a variety of body systems and a mix of medical or surgical treatments. Computer software was used with automated discharge abstract data to classify patients with any one of these 14 discharge diagnoses according to stage of disease and timeliness of hospital admission. Initial hospitalizations for the selected diagnoses totaled 1,549 for Hospital A and 1,164 for Hospital B.

A random sample of admissions to each hospital, stratified to assure representation of all 14 diagnoses and the three categories of timing of hospitalization, was selected for chart review to validate the diagnosis, staging, and hospitalization timing classifications assigned using the computer software. The sample included 532 admissions to Hospital A and 519 admissions to Hospital B.

The hospital medical records of these sample patients were abstracted using forms designed for each of the 14 sample diagnoses. Data were collected on patient demographic characteristics, admission and discharge diagnoses, preadmission medical history, laboratory evaluations, disease-specific complications, the presence of other diseases (comorbidities), and discharge status. The abstracted data were reviewed by physician consultants. Diagnosis, staging, and timing-of-hospitalization assignments were confirmed or, if necessary, changed on the basis of the physician review.

Using the sample data, a comparison was made between the combined mean lengths of stay for patients with one of the 14 selected diseases whose hospitalizations were early, timely, or late. Separate, within-hospital analyses were made between the mean lengths of stay for the timely and late categories. Comparisons were also made of hospital charges at Hospital A. (Analyses of hospital charges for patients in Hospital B were not performed because complete charge data were not available.) The *t*-test was performed to test the statistical significance of observed differences in the independent variables (timing-of-hospitalization categories) and in the means of the dependent variables (length of hospital stay and charges).

The full population data were used as the basis for an analysis of the rates of late hospital admissions for the 14 diagnoses and for an analysis of in-hospital mortality rates.

Results

Late hospitalizations for all patients with any one of the 14 diseases totaled 322 at Hospital A (21% of all patients treated for these diseases) and 231 at Hospital B (19% of all such patients).

The mean lengths of stay for timely and late hospitalizations in the stratified sample of the 14 diseases were compared for the two hospitals. For timely hospitalizations, the mean lengths of stay at Hospitals A and B were the same: 7.1 days. The mean lengths of stay for late hospitalizations were 18.2 days at Hospital A and 14.6 days at Hospital B.

The numbers of patients sampled from the timely and late groups, and the means and standard deviations of the lengths of stay at Hospital A, are presented in Table 1. For all 14 diseases the mean length of stay was longer for patients judged to be hospitalized late than for those whose hospitalizations were timely. Significant differences in lengths of stay were observed for five categories: bacterial pneumonia, cancer of the cervix, diabetes mellitus, diverticular disease, and urinary tract infection.

At Hospital A, five conditions, including external hernia, otitis media, pelvic inflammatory disease, purulent sinusitis, and uterine fibroma, had between 12 and 28 patients in the timely category but fewer than five patients in the late category. The small number of late hospitalizations makes statistical comparisons of mean lengths of stay between timely and late hospitalizations inappropriate. However, for these five diseases, the mean lengths of stay were also markedly longer for the late hospitalizations.

Combining all 14 conditions, the mean length of stay for late hospitalizations was 18.2 days, significantly exceeding the mean length of stay for timely hospitalizations, which was 7.1 days ($p < .01$).

The length-of-stay distributions of all timely and late hospitalizations at Hospital A for all 14 diseases combined are displayed in Figure 1. Less than 25% of the patients hospitalized on a timely basis had lengths of stay exceeding ten days, with very few patients (<10%) staying 20 days or longer. The length-of-stay pattern for the late hospitalizations was more highly skewed toward longer durations. Over half of these patients stayed more than ten days and over 20% stayed 20 days or longer.

Comparisons between mean lengths of hospital stay for the timely and late categories of the sample of patients who were admitted to Hospital B with any one of the 14 selected diseases are shown in Table 2. As in the analysis of mean lengths of stay for the patients at Hospital A, late hospitalizations were found to be consistently longer. Significant differ-

Table 1

Numbers of Patients, Mean Lengths of Stay (LOS), and *t*-Test Results for Timely and Late Hospitalizations in Hospital A, by Diagnostic Category, Fiscal Year 1984*

Diagnostic Category	Timeliness of Hospitalization						
	Timely			Late			
	n	Mean LOS	SD	*n*	Mean LOS	SD	*t*†
Appendicitis	21	9.6	15.4	16	12.9	8.5	n.s.**
Bacterial pneumonia	19	7.6	4.3	18	25.1	20.2	3.59¶
Cancer of cervix uteri	10	3.7	2.6	19	9.8	3.6	4.76¶
Cancer of colon and rectum	10	11.6	9.1	25	17.9	9.0	n.s.**
Cholecystitis	20	10.4	4.6	15	13.9	8.2	n.s.**
Diabetes mellitus	20	10.7	6.2	21	23.7	19.5	2.90¶‡
Diverticular disease	17	9.8	4.5	15	26.1	28.7	2.18‖§
External hernia	28	4.5	3.4	4	23.5	17.5	‡
Otitis media	19	3.5	3.2	2	8.5	7.8	‡
Pelvic inflammatory disease	12	4.6	1.5	0	—	—	‡
Peptic ulcer	22	6.9	5.3	20	11.2	9.2	n.s.**
Purulent sinusitis	25	5.3	5.3	2	53.5	.70	‡
Urinary tract infection	8	4.1	5.6	20	19.5	17.7	3.48¶§
Uterine fibroma	12	6.3	2.9	1	12.0	—	‡
All conditions	243	7.1	6.8	178	18.2	16.3	8.51¶§

*Over a one-year period, 1,549 patients admitted to Hospital A were classified according to whether their admissions were early, timely, or late, using criteria developed by a panel of physicians. This table compares data about a sample of the Hospital A patients whose admissions were timely with those of patients from the same hospital whose admissions were late, a total of 421 patients. The data show that the mean length of stay was longer for the patients who were hospitalized late, for all diseases. See Table 2 for comparable data from a sample of patients from Hospital B, the other hospital studied.

†The *t*-test for independent groups.

‡Insufficient sample size (fewer than five patients in one diagnostic group).

§Statistical test for homogeneity of variances was significant; therefore, *t*-test for unequal variances was applied.

‖$p<.05$

¶$p<.01$

**Not significant.

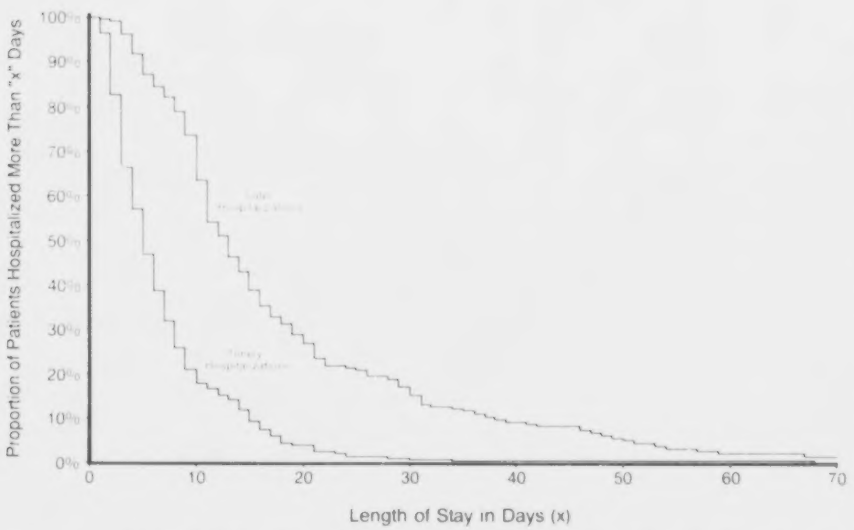


Figure 1. **The distribution of hospital lengths of stay for timely and late hospitalizations at Hospital A. The difference between the two curves indicates a higher probability of long stays for patients hospitalized late. Data from 532 admissions to Hospital A were used for this figure.**

ences between mean lengths of stay in the timely and late categories were observed for five of the diseases: appendicitis, bacterial pneumonia, cholecystitis, diverticular disease, and urinary tract infection.

As in the analysis of Hospital A data, small numbers of patients were found in the late-hospitalization group for six of the diseases, obviating statistical analysis. For each of these diseases — cancer of the cervix uteri, external hernia, otitis media, pelvic inflammatory disease, purulent sinusitis, and uterine fibroma — fewer than five patients had late hospitalizations. When mean length-of-stay data were determined for all 14 diseases combined at Hospital B, a significantly greater mean length of stay of 14.6 days was found for patients in the late-hospitalization category compared with 7.1 days for patients in the timely category ($p < .01$).

Table 2

Numbers of Patients, Mean Lengths of Stay (LOS), and *t*-Test Results for Timely and Late Hospitalizations in Hospital B, by Diagnostic Category, Fiscal Year 1984*

Diagnostic Category	Timeliness of Hospitalization: Timely			Late			
	n	Mean LOS	SD	*n*	Mean LOS	SD	*t*†
Appendicitis	24	3.2	1.0	16	8.9	3.3	6.89¶§
Bacterial pneumonia	19	8.2	4.9	20	17.0	14.8	2.70‖§
Cancer of cervix uteri	6	2.0	2.4	1	6.0	—	‡
Cancer of colon and rectum	17	19.9	12.3	23	20.0	12.6	n.s.**
Cholecystitis	22	8.6	2.8	19	13.6	8.0	2.59¶§
Diabetes mellitus	21	7.7	3.9	18	12.2	9.3	n.s.**
Diverticular disease	15	9.6	4.6	6	25.8	14.6	2.66‖§
External hernia	21	3.6	2.2	4	29.8	18.3	‡
Otitis media	11	4.3	3.0	2	5.5	.7	‡
Pelvic inflammatory disease	20	6.2	4.2	1	4.0	—	‡
Peptic ulcer	19	8.2	4.7	19	12.5	8.5	n.s.**
Purulent sinusitis	14	3.9	4.6	0	—	—	‡
Urinary tract infection	20	5.8	4.9	23	11.7	4.6	4.04¶
Uterine fibroma	18	6.5	2.5	0	—	—	‡
All conditions	247	7.1	6.2	152	14.6	10.9	7.70¶

*Over a one-year period, 1,164 patients admitted to Hospital B were classified according to whether their admissions were early, timely, or late, using criteria developed by a panel of physicians. This table compares data about a sample of Hospital B patients whose admissions were timely with those of patients from the same hospital whose admissions were late, a total of 399 patients. The data show that the mean length of stay was longer for the patients who were hospitalized late, for most diseases. See Table 1 for comparable data from a sample of patients from Hospital A, the other hospital studied.

†The *t*-test for independent groups.

‡Insufficient sample size (fewer than five patients in one diagnostic group).

§Statistical test for homogeneity of variances was significant; therefore *t*-test for unequal variances was applied.

‖$p<.05$.

¶$p<.01$.

**Not significant.

As displayed in Table 3, the mean charges for the Hospital A patients in the late category exceeded the mean charges for the same hospital's patients in the timely group for all diseases. For all conditions combined, the mean charge was $6,821 for patients hospitalized on a timely basis, compared with $20,728 for patients hospitalized late. (Hospital B data for these types of charges were not available.)

In eight of the diagnostic categories, and for all conditions combined, mean charges for late hospitalizations were significantly greater than mean charges for timely hospitalizations ($p < .05$). For three of these eight conditions, cancer of the colon and rectum, cholecystitis, and peptic ulcer, significant differences were observed in mean charges despite the fact that observed differences in mean lengths of stay were not significant for these diseases. Five of the diagnostic categories had too few late hospitalizations to justify a statistical test of significance.

An analysis of timely and late hospitalizations of two diseases, bacterial pneumonia and cholecystitis, offers examples of the wide range of hospital resource use within single diagnosis-related groups (DRGs). Eighteen patients with bacterial pneumonia in the sample population were grouped into DRG 89: "simple pneumonia and pleurisy, age greater than or equal to 70 and/or complicating conditions." When these patients were divided by timeliness of hospitalization, marked differences were evident in mean lengths of stay and mean charges. Eight patients whose hospitalizations were categorized as timely stayed 9.5 (SD 4.2) days, and their mean charge was $7,564 (SD $3,500); ten patients who were hospitalized late stayed 23.7 (SD 24.2) days, with a mean charge of $35,469 (SD $42,426).

Sixteen patients in the study population with cholecystitis were classified within DRG 197: "total cholecystectomy without common duct exploration, age greater than or equal to 70 and/or complicating conditions." Ten patients in the timely group stayed a mean of 9.8 (SD 3.9) days, with a mean total hospital charge of $9,549 (SD $3,326), in contrast to 13.3 (SD 11.4) hospital days with a mean charge of $15,602 (SD $10,646) for six patients in the late group.

Based on population data, the in-hospital mortality of all patients with a principal diagnosis of bacterial pneumonia shows an expected rise in mortality of the patients hospitalized in the late stage of the disease. At Hospital A, five of 132 patients

Table 3

Number of Patients, Mean Charges for Stay, and *t*-Test Results for Timely and Late Hospitalizations in Hospital A, by Diagnostic Category, Fiscal Year 1984*

Diagnostic Category	Timeliness of Hospitalization: Timely			Timeliness of Hospitalization: Late			
	n	Mean Charges in $	SD	*n*	Mean Charges in $	SD	*t*†
Appendicitis	21	10,505	16,294	16	13,494	9,058	n.s.**
Bacterial pneumonia	19	6,546	5,398	18	38,533	35,087	3.82¶§
Cancer of cervix uteri	10	3,902	2,514	19	11,532	4,772	4.70¶§
Cancer of colon and rectum	10	10,341	7,700	25	18,652	8,843	2.60‖
Cholecystitis	20	10,415	4,849	15	15,649	7,993	2.25‖§
Diabetes mellitus	20	8,391	5,018	20	22,640	24,625	2.54‖§
Diverticular disease	17	8,971	4,967	15	31,901	38,135	2.31‖§
External hernia	28	4,202	2,208	4	23,097	15,554	‡
Otitis media	19	4,335	3,190	2	6,777	3,380	‡
Pelvic inflammatory disease	12	4,844	1,278	0	—	—	‡
Peptic ulcer	21	6,412	5,615	20	16,179	19,653	2.14‖§
Purulent sinusitis	25	5,284	4,613	1	70,621	—	‡
Urinary tract infection	7	3,795	3,690	20	18,687	19,014	3.33¶§
Uterine fibroma	11	6,312	2,803	1	13,081	—	‡
All conditions	240	6,821	6,705	176	20,728	22,195	8.05¶§

*This table compares the hospitalization charges of a sample of Hospital A patients whose admissions were timely with the charges of patients from the same hospital whose admissions were late, a total of 416 patients. The data show that the mean charges for the patients in the late category exceeded the mean charges for those in the timely group, for all diseases. Comparable data on charges for patients in these categories from Hospital B, the other hospital studied, were not available.

†The *t*-test for independent groups.

‡Insufficient sample size (fewer than five patients in one diagnostic group).

§Statistical test for homogeneity of variances was significant; therefore, *t*-test for unequal variances was applied.

‖$p<.05$.

¶$p<.01$.

**Not significant.

(3.8%) who were categorized as having timely hospitalizations died, whereas 16 of 41 patients (39.0%) in the late category died. Similar proportions were observed in the same categories for patients in Hospital B: three of 144 patients (2.1%) in the timely category died, whereas 11 of 38 (28.9%) in the late category died. (For both hospitals, $p < .001$.)

Discussion

With the current emphasis on cost containment, it is increasingly important to gain greater understanding of the factors affecting the cost of health care services. The principal aim of utilization review programs is to assure that patients using an acute care hospital are "sick enough" to justify the use of that expensive resource. The prospective payment system is designed to provide financial incentives for the efficient use of those resources through payment of fixed fees for treatment of predefined diagnostic types. The authors believe that there is a major gap in the focus of these programs. Neither traditional utilization review programs nor the prospective payment system adequately accounts for the patient who arrives at the hospital "sicker"—at a later stage—than is appropriate.

In this project, disease staging was applied to define objectively the severities of patients' diseases, in order to create categories of "isomorbid" patients. The case-mix classification system can also be differentiated in a way that will allow assessment of the appropriateness of the timing of a patient's hospitalization. Using the instruments developed for this project, data were collected for a sample of patients who had any one of 14 diagnostic categories from two hospitals to test the principal hypothesis: Significantly fewer resources are consumed by a patient hospitalized on a timely basis than are consumed by a patient hospitalized later than desirable in the disease process.

The data from this study's two sample hospitals strongly support this hypothesis. The differences between timely and late admissions in mean lengths of stay and mean charges (from the hospital for which charge data were available) were striking. For every diagnosis for which sufficient data were available to make meaningful comparison, the mean length of stay and mean charge for late hospitalizations exceeded the mean length of stay and mean charge for timely hospitalizations.

For all of the sample diagnoses combined, the mean length of stay for late hospitalizations exceeded the mean length of stay for timely hospitalizations by 11.1 days at one hospital and by 7.5 days at the other hospital. The mean charge per late

hospitalization exceeded the mean charge per timely hospitalization by $13,907. For the 176 late hospitalizations included in the sample at hospital A, this differential amounted to over $2.4 million. This difference in charges resulted from the greater intensity of services (diagnostic and/or therapeutic) needed for these patients.

While the project did not focus on comparisons between the two sample hospitals, the proportions of late hospitalizations and the pattern of length of stay were quite similar in the two hospitals, indicating that the problem of late hospitalizations is not unique to large academic medical centers.

The intra-DRG analysis of bacterial pneumonia and cholecystitis patients highlights the heterogeneity found in a number of DRGs. Under a prospective payment system using DRGs for definition of case type, groups of patients who are consistently hospitalized late can be very costly to the hospital.

Insofar as late-stage treatment is costly and leads to increased risk of morbidity and mortality, it is important that factors related to delayed treatment be identified so that changes may be implemented. If a patient is treated only at a late stage of illness in a hospital, something has gone wrong. Possible explanations include

1. The disease has been treated appropriately on an outpatient basis but a lack of medical knowledge prevents its successful treatment.
2. The patient has not received appropriate medical treatment on an outpatient basis because of diagnostic or other physician errors.
3. The patient has not complied with treatment provided by the physician.
4. The patient did not seek early medical care for any of a variety of reasons, such as fear, ignorance, and lack of access to care.
5. The disease progression was so rapid or unobservable that the patient would have been unlikely or unable to seek early medical treatment.
6. The patient was hospitalized in a timely fashion but inpatient treatment was delayed or inappropriate.

Undoubtedly not all late hospitalizations observed in this study were preventable; however, hospitalizing even a small proportion of these patients at an earlier stage could have yielded substantial cost savings, as well as potential reductions in morbidity and mortality.

The techniques described in this paper can be used to identify patterns of inappropriately late hospitalization. Where such patterns are found, corrective action(s) focused on the identified problem areas can be implemented. Depending upon the contributions of the patient, the physician, the health care delivery system, and the community to these problems, appropriate activities might include specific patient education programs, patient screening programs, continuing medical education programs, and feedback to physicians and organizations that provide health care.[11]

Measurement of timing of hospitalization can also be used as a tool for the evaluation of the outcome of care provided in a variety of forms of outpatient programs, such as health maintenance organizations (HMOs). Theoretically, HMOs and other managed care organizations, through improved accessibility and continuity of care, should have proportionately fewer inappropriately early and late hospitalizations than would a fee-for-service system. Evaluating the patterns of timing of hospitalization of individuals enrolled in these types of plans in comparison with similar groups of patients would enable the testing of this type of hypothesis.

The dramatic changes in the organization and financing of the health care system necessitate the development and implementation of tools that can be used on a large scale to evaluate their impact on quality of care. The authors believe that the approach described in this paper can be used as a part of such evaluations.

The investigation reported in this article was supported by a grant from the W. K. Kellogg Foundation, Battle Creek, Michigan.

References

1. Siu, A. L., et al. Inappropriate Use of Hospitals in a Randomized Trial of Health Insurance Plans. *N. Engl. J. Med.* **315**(1986):1259-1266.
2. Sandrick, K. M. Blue Cross and Blue Shield of Michigan's Efforts to Change Practice Patterns. *ORB* **10**(1984):349-352.
3. Shahoda T. Preadmission Review Cuts Hospital Use. *Hospitals* **58**(August 1984):54-55.
4. Social Security Amendments of 1983, Public Law 98-21, 97 Stat., 65-172.
5. Schlesinger, M. On the Limits of Expanding Health Care Reform: Chronic Care in Prepaid Settings. *Milbank Mem. Fund Q.* **64**(1986):189-214.
6. *Report to Congress: Impact of the Medicare Hospital Prospective Payment System, 1984 Annual Report.* Department of Health and Human Services, Health Care Financing Administration, Office of Research and Demonstration, HCFA Pub. No. 03231. Washington, D.C.: Government Printing Office, 1984.
7. Gonnella, J. S., Hornbrook, M. C., and Louis, D. Z. Staging of Disease: A Case-Mix Measurement. *JAMA* **251**(1984): 637-644.
8. Gonnella, J. S., Louis, D. Z., and McCord, J. J. The Staging Concept: An Approach to the Assessment of Outcome of Ambulatory Care. *Med. Care* **14**(1976):13-21.
9. Louis, D. Z., Barnes, C. A., and Pepitone, T. *Disease Staging: A Clinically Based Approach to Measurement of Disease Severity.* Vols. 1-5, (PB83-254631). Rockville, Maryland: National Center for Health Services Research, 1983.
10. Conklin, J. E., Lieberman, J. V., Barnes, C. A., and Louis, D. Z. Disease Staging: Implications for Hospital Reimbursement and Management. *Health Care Financ. Rev. Supp.* [no volume no.] (November 1984):13-22.
11. Gonnella, J. S., Zeleznik, C. Factors Involved in Comprehensive Patient Care Evaluation. *Med. Care* **12**(1974):928-934.

Evaluation of Students in Medicine Clerkships

GREGORY J. MAGARIAN, M.D., and DENNIS J. MAZUR, M.D., Ph.D.

Abstract—This article reports the results of a survey of the 124 medical schools in the United States and Puerto Rico, mailed in late 1986; the majority of the responses from 101 schools were received in early 1987. The authors report their findings concerning the overall process used by these schools to evaluate medicine clerks, specifically identifying the importance given to subjective evaluations, made by attending physicians and/or housestaff, compared with the importance given to clerks' performances on objective means of evaluation based on written and/or oral examinations, whether national or faculty-developed. The study reveals that the subjective evaluations of clerks by faculty attending physicians and by housestaff working with the clerks were used by 98 and 83 of the responding schools, respectively, and were given the greatest evaluative weight. Examinations of all types were used by 87 of the responding schools. The National Board of Medical Examiners (NBME) Part II Medicine Examination was the most frequently used and heavily weighted examination, followed by written faculty examinations. Oral examinations were the least frequently used evaluative tools, carried the least weight, and were rarely used without another examination. In light of their findings, the authors maintain that further research should be done to clarify whether the NBME Part II Medicine Examination is an appropriate instrument to measure the knowledge acquired during clerkship experiences. *Acad. Med.* 65(1990):341–345.

Students enter their clinical years anxious about leaving the classroom environment, where the guidelines regarding the evaluation process have been well established. For the most part, success during the basic science years is measured by performance on examinations. Personal and professional attributes, ability to work well within a team, and the ability to work closely with patients and others are not necessarily significant factors in the evaluation of students during these years. However, these attributes are critical to success as a ward clerk, houseofficer, and physician. The ability to organize responsibilities to patients and other team members must be balanced against developing skill in patient write-ups for review and case presentations to attending physicians. Throughout their clinical years, students are subject to evaluations that are likely to be far more subjective than what they have previously experienced. Of great concern to students is that evaluations be fair and consistent and that they know what is expected for achieving different levels of performance.

This study was one component of a larger study based on a survey of medicine clerkship directors. Here we report the answers to four questions concerning the overall process of evaluating clerks as reported by medicine clerkship directors in the United States and Puerto Rico: (1) What role do subjective evaluations of attending physicians and/or housestaff play in the evaluation of medicine clerks? (2) What role does the performance of medicine clerks on written and oral examinations play in this evaluation process? (3) What types of examinations are used in this evaluative process? (4) What sources of guidance, if any, are given to medicine clerkship directors in determining how to evaluate medicine clerks?

Method

The authors surveyed 124 medical school medicine clerkship directors in the United States and Puerto Rico in the latter part of 1986; responses were received in early 1987. The questionnaire asked the directors to record the percentages of weight given to the following subjective and/or objective components used in grading medicine clerks: (1) subjective evaluations from faculty serving as attending physicians, (2) subjective evaluations from housestaff working directly with students, (3) the National Board of Medical Examiners Part II Medicine Examination (subsequently referred to as the NBME–med exam), (4) written examinations developed by faculty, and (5) oral examinations.

Results

Responses were received from 101 (85%) of the 124 medical schools surveyed. Table 1 lists the components and their relative weights used in grading reported by these schools, in descending order of the number of schools using each component. The five most frequently used components were those offered as choices on the questionnaire; the remainder were spontaneously reported by the schools. When a single percentage was given to estimate the weight of evaluations received from both attending physicians and housestaff, the authors gave each of these evaluation categories equal weight (that is, each was assigned half the total percentage that had been given the combined category). This assignment of equal weight was made for responses from 17 of the schools.

Table 2 lists the five major components from which grades were derived and categorizes each into percentiles of distribution. Lack of attending physician faculty input and lack of housestaff input into a student's final evaluation were reported by three and 19 schools, respectively. This table shows that attending physician faculty and housestaff evaluations carried the greatest weight in the evaluation of medicine clerks, followed

Dr. Magarian is director of the medicine clerkship and associate professor, and Dr. Mazur is associate director of the medicine clerkship and assistant professor; both are in the Department of Medicine, Oregon Health Sciences University, Portland. Both authors are also staff physicians in the Section of General Internal Medicine, the Portland Veterans Administration Medical Center, Portland, Oregon.

Correspondence and requests for reprints should be addressed to Dr. Magarian, VA Medical Center (111-P), 3710 SW U.S. Veterans Hospital Road, P.O. Box 1034, Portland, OR 97207.

Table 1

Averages and Ranges of Weights That 101 U.S. Medical Schools Assigned to 13 Components Used in Determining the Final Grades of Medicine Clerks, and the Number of Schools Using Each Component, 1986*

Component Used in Grading Clerks	Average of Weights Given by All Schools (%)†	Range of Weights (%)	No. Schools Using Component
Attending physicians' evaluations	42	10–100	98
Housestaff evaluations	33	10–85	82
National Board of Medical Examiners Part II Medicine Examination‡	27	10–65	57
Written faculty-developed examinations	25	5–50	40
Oral examinations	15	5–30	26
Chief of medicine evaluations	12	10–15	3
Objective Structured Clinical Examination	33	25–40	2
Tutors' evaluations	40		1
Practical clinical examinations	25		1
Professional habits and attitudes	25		1
Students evaluating each other	15		1
Written patient work-ups	15		1
Clinic performance	10		1

*The authors surveyed 124 medical schools in the United States and Puerto Rico in late 1986 concerning the importance given to (1) subjective and (2) objective (written and/or oral) means of evaluating medicine clerks. A total of 101 schools (80%) responded.

†When a single percentage was given to estimate the weight of evaluations received from both attending physicians and housestaff, the authors gave each category equal weight (that is, each was assigned half the total percentage that had been given the combined category). This assignment of equal weight was made for 17 of the responses.

‡One school used this component only to elevate the final grade when there was superior performance on the examination.

closely by the use of examinations. There was a direct correlation between the use of the five major components used in grading and the average weight given to that component of the grade. In other words, the frequency of use and the weight given to a specific evaluation component were both in descending order: attending physician faculty evaluations, housestaff evaluations, NBME–med exam performance, written examination performance, and oral examination performance. Other evaluative methods were reported very infrequently.

Table 3 reveals how frequently each of the five major components were used in various combinations in the evaluation of students. The most common means of evaluating students was the use of input from attending physician faculty and/or house officers who had worked with students during the clerkship (98 schools). Only 12 schools reported that their evaluations of student performance were exclusively from attending physician faculty and housestaff input, without use of any examination. In only one school was the entire evaluation and grade determined solely from faculty attending input. Seventeen (17%) of the 98 schools using faculty attending physicians' input in their evaluation processes did not use housestaff input in this process, whereas only one of the 82 schools using housestaff input in the process did not also use faculty evaluations.

A total of 87 of the 101 responding schools reported using one or more written examinations of some type. The NBME–med exam was used by 57 of these 87 schools (66%); faculty-developed written examinations were used by 40 schools (46%); and oral examinations administered by faculty were used by 26 schools (30%). A total of 55 of the 87 schools that reported using examinations used only one type of examination (63%). The NBME–med exam was used as the single examination by 29 of these 55 schools (53%); faculty-developed examinations were used in this way by 23 schools (42%); and oral examinations administered by faculty were used as the single examination by only three schools (5%). The Objective Structured Clinical Examination (OSCE) was used as the single examination in only one of the 87 schools.

The frequencies of the various combinations used to evaluate medicine clerks were reported as follows: overall, only one examination was used more frequently than were two or more (55 of the 87 schools using examinations, or 63%, versus 32 of these schools, or 37%). Of the 32 schools using two or more examinations, 15 (47%) used the NBME–med exam plus an oral examination; nine (28%) used the NBME–med exam plus a faculty-developed written examination; four (13%) used a faculty-developed written examination plus an oral examination; and four (13%) used all three of the just-mentioned examination types. In only four (13%) of the 32 schools using two or more examinations was the NBME–med exam not included.

Only two schools reported using the OSCE, although two others specifically noted plans to institute it during the following academic year, while yet others were considering it. Departmental chairpersons provided input that was assigned a specific percentage of a student's final grade in only three schools. It is interesting that in one school, students' evaluations of each other accounted for 15% of the medicine clerkship grade. Only one school reported deriving a specified percentage of students' grades from students' professional habits and attitudes (25%), while one other reported giving a specific weight (10%) to students' performances in outpatient clinics.

A single examination was given greater weight than cumulative faculty attending evaluations in 20 schools and greater weight than cumulative housestaff evaluations in 29

schools. Equal weight was given to examinations compared with cumulative faculty attending physician and housestaff evaluations of clerks in 13 and 14 schools, respectively. In both instances, the NBME-med exam was most commonly used, followed in descending order by faculty-developed written examinations and oral examinations. Overall, an examination carried greater or equal weight compared with cumulative evaluations from attending physician faculty or house officers in a third (34) of the responding schools. However, in two-thirds (67 of 101) of the responding schools, subjective assessments by faculty or housestaff working with clerks carried the greatest weight in determining final grades.

Review of how results from the NBME-med exam were used revealed no consistent means of using the information in assisting in the evaluation and grading of students. Fifty schools used the raw scores as the primary means of evaluating student examination performance. Typically, specific raw scores were established to separate different levels of performance, such as distinguishing honors from passing. The remainder of the schools used a variety of other means to interpret NBME-med exam performances by medicine clerks. In rank order, these included performance within the class or based on data from prior classes, performance based on the national percentiles of previous fourth-year students having taken the examination, a combination of raw score and performance within the class, a combination of national percentile and performance within the class, and (as reported by one school) a combination of raw score and national percentile.

Only 32 of the 101 responding clerkship directors reported that they had received any guidance in how to evaluate students, while 81 schools reported that faculty had received some form of instruction in the evaluation process. This was most commonly from memoranda and departmental meetings.

Table 2

Numbers of U.S. Medical Schools That Assigned Specific Weights to Five Major Components Used in Determining the Final Grades of Medicine Clerks, Listed by Percentile Distribution, 1986*

	Component Used in Grading Clerks				
Percentile Distribution of Weights	Evaluations of Attending Physician No. Schools	Evaluations of Housestaff No. Schools	NBME-med Exam† No. Schools	Written Exams No. Schools	Oral Exams No. Schools
0-10	2	5	2	8	12
11-20	4	13	15	10	6
21-30	22	22	22	10	6
31-40	30	23	9	7	—
41-50	21	12	4	4	—
51-60	6	—	—	—	—
61-70	1	2	—	—	—
71-80	1	1	—	—	—
81-90	1	1	—	—	—
91-100	1	—	—	—	—
Unknown	9	3	5	1	2
Total	98	82	57	40	26

*The authors surveyed 124 medical schools in the United States and Puerto Rico in late 1986 concerning the importance given to (1) subjective and (2) objective (written and/or oral) means of evaluating medicine clerks. A total of 101 schools (80%) responded.

†National Board of Medical Examiners Part II Medicine Examination.

Table 3

Numbers of 101 U.S. Medical Schools Using Specified Combinations of Components to Determine the Final Grades of Medicine Clerks, 1986*

Combination of Components Used in Grading Clerks†	No. Schools Using Combination
Attending physician, houseofficer, NBME-med exam	22
Attending physician, houseofficer, written exam	19
Attending physician, houseofficer, NBME-med exam, oral exam	12
Attending physician, houseofficer	11
Attending physician, houseofficer, NBME-med exam, written exam	7
Attending physician, NBME-med exam	7
Attending physician, houseofficer, written exam, oral exam	4
Attending physician, written exam	4
Attending physician, houseofficer, oral exam	3
Attending physician, NBME-med exam, written exam	2
Attending physician, NBME-med exam, written exam, oral exam	2
Attending physician, NBME-med exam, oral exam	2
All other combinations each used by only one school or indeterminable (one school)	6

*The authors surveyed 124 medical schools in the United States and Puerto Rico in late 1986 concerning the importance given to (1) subjective and (2) objective (written and/or oral) means of evaluating medicine clerks. A total of 101 schools (80%) responded.

†NBME-med exam = National Board of Medical Examiners Part II Medicine Examination.

Discussion

This study of the components used by departments of medicine and their clerkship directors in the evaluation of medicine clerks has revealed that the evaluation of medicine clerks is heavily based on subjective input provided by faculty and housestaff phy-

sicians. It is not known how variable or consistent the overall processes of evaluation by faculty and housestaff are. Faculty and housestaff will have worked for variable periods of time with medical students, typically on hospital inpatient services. Only one school reported the use of student performance in an outpatient clinic setting as a specific component of its overall evaluation of students.

The subjective evaluations of attending physicians and those of housestaff were the foundation of the evaluation process, being used by 98 and 82 schools, respectively. While 17 of 98 schools used faculty attending evaluations without evaluations from housestaff working directly with students, only one school reported using evaluations from housestaff without using evaluations by faculty in the evaluation of student performance on medicine clerkships. Furthermore, in only five schools of 81 reporting use of evaluations from both sources were the housestaff evaluations clearly weighted heavier than those from faculty, whereas the reverse was true in 22 schools. Therefore, overall, evaluations from attending physicians were the most frequently used and the most heavily weighted. These evaluations were also the most frequently used when only one evaluation component was used.

Because of the weight given to such input and the differences in backgrounds and the variable amounts of evaluative experiences of individual faculty members and housestaff, it is clear that all faculty and housestaff evaluating students should be well versed in departmental expectations and trained in proficiency in the evaluation process. In this regard, it is heartening that 81 medicine clerkship directors reported that their faculties had received some form of guidance in the student evaluation process. However, the authors of the present study believe memoranda or discussions in departmental meetings, the primary sources reported for such guidance, may be inadequate, and that workshops,[1] as reported by only 13 schools, may be far more effective, despite their labor intensity and potential expense. While one school reported the development of a new medicine clerkship curriculum to assure high-quality teaching, independent of teaching by housestaff and by attending physicians,[2] the authors do not believe it is feasible for most schools to develop an additional evaluative staff of housestaff and attending physicians to evaluate medicine clerks. This should be an endeavor shared by all attending physicians and housestaff working with clerks in inpatient or ambulatory care arenas.

Performance on examinations is also a major indicator of achievement used in the evaluation of medicine clerks. Others[3-5] have noted that clinical ratings alone are not adequate for measuring student progress on medicine clerkships. They argue that a comprehensive assessment of student performance in medicine clerkships should include written examinations as well as clinical ratings. Only 14 schools did not report using examinations of some form in their evaluations of clerks. For the three examinations used most often, the frequencies of their use reflected the relative weights applied to them.

The NBME-med exam was the examination most frequently used and most heavily weighted in the evaluation process according to responses from medicine clerkship directors in this survey. This may reflect less confidence in faculty-developed examinations and/or the need for more effort in their creation. However, despite its frequent use for evaluating medicine clerks, the NBME-med exam was not created for this purpose. Furthermore, to the authors' knowledge, no study has demonstrated its appropriateness in measuring clinical knowledge and judgment acquired during medicine clerkships. Performance on this examination at the end of medicine clerkships may reflect the overall accumulation of clinical knowledge and judgment up to that point in time in a student's total medical school experience, including other preceding clerkships. If such is the case, one might expect students to enter with a greater baseline medical knowledge base and perform better on the NBME-med exam when the medicine clerkship is taken later in the year, after they have had more clinical experiences from other preceding clerkships. If so, this needs to be taken into account to assure fairness in grading. The popularity of the NBME-med exam probably also relates to its reasonable cost ($25 per examination), ease of administration, development and standardization by professional test makers, and national normative values for comparing student performance.

Although over half of the responding schools used the NBME-med exam for clerkship testing, the normative values returned by the NBME did not include the one that would be most valuable, namely, how a school's medicine clerks performed in comparison with the medicine clerks from other schools taking the examination during the same period. Rather, the NBME-med exam provided normative data to compare a school's third-year clerks with a group of students who previously took the same examination when administered during their fourth year as part of the NBME examination—Part II. This is likely to have contributed greatly to the marked variability in the way individual schools used the results in their overall evaluation processes. For the NBME-med exam to continue to enjoy such popularity for use as an examination in medicine clerkships, its validity as an appropriate tool in this regard needs to be established. Although the weight given to performance on this examination is higher than that given to faculty-developed written or oral examinations, justification for this test's elevated stature has not been established. The authors are unaware of data on the correlation between performances of medicine clerks on the NBME-med exam and concomitant performances on written or oral examinations. A previous study[4] revealed poor correlation between NBME-med exam performance and the evaluations of faculty who had worked with the same students, and suggested that a halo effect may be playing a role. This latter effect would not be expected to affect written examinations or oral

examinations, if conducted by examiners unaware of a student's performance as perceived by others.

Oral examinations were used as the only examination by only three schools (12% of schools using single examinations), which probably reflects their labor intensity and an overall distrust of their use as a consistent and fair measure of students' knowledge and judgment. In only four schools were oral examinations paired with another faculty-developed written examination. In other words, when oral examinations were used, 70% of the time they were used in conjunction with the standardized NBME-med exam.

Summary

The results of this survey strongly suggest that faculty physicians and housestaff are the most heavily relied upon to evaluate medicine clerks. It would seem important to assure that they are adequately trained to do so beyond simply having received guidelines from memoranda and discussions in departmental meetings. Workshops may be particularly effective in this regard.

Likewise, if the NBME-med exam is to be used in clerkship evaluations, it is important to establish whether it measures what is wanted. When taken by most students in their fourth year, it provides a reflection of where a student is, relative to his peers, locally and nationally, at that point in time, based on multiple clinical experiences beyond the medicine clerkship. The question remains: does performance on the NBME-med exam provide a useful indicator of knowledge acquired during the clerkship, or does it primarily provide a measure of a student's knowledge based on all his or her medical school experiences from both basic science and clinical studies? If the latter is true, use of the NBME-med exam may not be appropriate as a tool for evaluating the development of students during medicine clerkships. Similarly, if faculty-developed written or oral examinations are used, training in their development and use would seem equally important, to assure that they also measure what is desired. The authors believe a major factor leading to the relative lack of use of faculty-developed examinations is that faculty are uncertain of their skills in the development of such examinations and of whether such examinations are fair and consistent. This uncertainty may be a reflection of the paucity of clerkship directors who have received specific training in the evaluation of clerks. The authors believe that for improvement to occur in the overall evaluation process of students on medicine clerkships, it is important that faculty and housestaff receive more and better training in the evaluation process. However, since the minority of medicine clerkship directors reported having such training themselves, it is not clear how and by whom this training would occur and who (medical schools versus individual departments) would fund it.

References

1. Vadiserri, R. O., Koziol, S. M., Korth, W. W., and Haynes, L. T. A Workshop to Improve the Teaching Skills of Physician Faculty Members. *J. Med. Educ.* **61**(1986):469-471.
2. Regan-Smith, M. G. Improved Teaching Within a Medicine Clerkship. *J. Med. Educ.* **59**(1984):753-755.
3. Ramsey, P. G., et al . Use of Objective Examinations in Medicine Clerkships: Ten-Year Experience. *Am J. Med.* **81**(1986):669-674.
4. Marienfeld, R. D., and Reid, J. C. Subjective vs. Objective Evaluation of Clinical Clerks. *N. Engl. J. Med.* **302**(1980):1036-1037.
5. O'Donohue, W. J., Jr., and Wergin, J. F. Evaluation of Medical Students During a Clinical Clerkship in Internal Medicine. *J. Med. Educ.* **53**(1978):55-58.

(WONCA).

This new model has had a positive impact in Malta. Through WONCA, the model has attracted the attention of China and several countries in Latin America. In China it was discussed at the Chinese Medical Association's First International Scientific Meeting of General Practice in November 1989.

Reference

1. College of Family Physicians of Canada. *Educational Objectives for Certification in Family Medicine.* Toronto, Canada: CFPC, 1981.

SHORT PAPER

The Availability of Fellowship Training for Foreign Family Physicians

MICHAEL D. COHEN, M.B., Ch.B., and JOEL H. MERENSTEIN, M.D.

Abstract—The authors surveyed the 49 U.S. fellowship programs in family medicine listed in the 1988 *Fellowship Directory* published by the Society of Teachers of Family Medicine. The past experience and future intentions of the 47 responding programs toward non-immigrant foreign family physicians coming to the United States to undergo fellowship training in academic family practice were studied. Only a minority of the programs (22%) had ever trained a foreign fellow, and of these only five of 11 were definitely able to do so in the future. Over half (52%) of the programs that had never trained a foreign fellow stated that they were not prepared to train a foreign physician. Although foreign family physicians are encouraged by the American Academy of Family Practice to come to the United States for fellowships, this study shows that it is extremely difficult for these physicians to find programs willing to act as host. *Acad. Med.* 65(1990):348–350.

Although fellowships in academic family practice exist in the United States, they are uncommon in other countries. Letters written to the representative family practice organizations in seven English-speaking countries of the world (Australia, Britain, Canada, Ireland, New Zealand, South Africa, and the United States) revealed that there is no directory of institutionally funded academic family practice fellowships outside the United States and that few countries have any organized program to prepare family physicians for a career in academic family practice.

Apparently aware of this, the American Academy of Family Physicians has established a policy to encourage foreign physicians to come to the United States to study family practice and return to promote it in their home countries.[1]

The purpose of this study was to evaluate the experience and availability of fellowship training for foreign family physicians in the United States.

Method

The authors sent a letter to each of the fellowship programs listed in the 1988 *Fellowship Directory* published by the Society of Teachers of Family Medicine[2] asking whether they had trained any foreign fellows in their departments. A stamped, self-addressed postcard was provided for the reply. A foreign fellow was defined as a physician who fulfilled the following criteria: (1) was born outside the United States and is not a U.S. citizen; (2) has not studied at a medical school in the United States; (3) has undergone residency training in family practice outside the United States; (4) has undergone a fellowship program in academic family practice in the United States of at least one

Dr. Cohen is director of Be'ere Health Center of Kupat Cholim, Netanya, Israel. At the time of this study, he was a fellow in family practice at St. Margaret Memorial Hospital, Pittsburgh, Pennsylvania. Dr. Merenstein is director of research and fellowships in the family practice residency, also at St. Margaret Memorial Hospital.

Correspondence and requests for reprints should be addressed to Dr. Cohen, Director, Be'ere Clinic, Be'ere Street, Netanya, Israel.

A related article, "A New Faculty Training Model for Countries Developing Academic Family Medicine Programs," appears on page 346.

year's duration; and (5) intends to return to his or her home country at the end of the fellowship and has not come to the United States as an immigrant.

When a reply was received, a second, more detailed questionnaire was sent. If the program had trained a foreign fellow in the past then the questionnaire asked whether the organizers intended to train foreign fellows in the future. A second questionnaire was also included concerning each of the fellows who had been or were being trained, asking specific questions concerning the fellow's country of origin and other aspects of the person's history to check whether the criteria for a foreign fellow, listed in the previous paragraph, had been met. If the program had not trained a foreign fellow in the past, then the questionnaire asked whether it was possible for foreign physicians to pursue fellowships at that institution and, if not, what the reasons for excluding them were. If a program did not reply at either stage of the survey, then, after a month had elapsed, the appropriate questionnaire was sent again. If the second questionnaire mailing did not evoke a response, then the program director was telephoned personally and the questions asked directly.

Many of the fellowship programs listed in the *Fellowship Directory* were no longer in existence and the correspondents and telephone numbers for some others were often quite different from the ones given in the directory.

Results

Initial responses were received from 47 of the 49 programs listed. Of these, 36 (77%) had never had a foreign fellow participate in their programs. For the second questionnaire a total of 46 programs responded: ten of the 11 programs that had had foreign fellows, and all of the 36 who had not trained foreign fellows.

These data show that only five of the ten programs that had trained foreign fellows in the past were sure that they would be able to do so in the future. Of the 36 programs that did not have a foreign fellow, ten would be prepared to take foreign fellows in the future. The largest number of programs had never trained a foreign fellow and did not intend to do so in the future. Of the 46 programs responding, 21 said they might consider a foreign fellow; of these 21, six were uncertain of having funding available for foreign fellows.

The reasons programs gave for not considering foreign applicants were: (1) that a medical license would be unobtainable (three programs); (2) that there could be potential linguistic and communications problems (one); (3) that the program no longer existed (three); (4) that American candidates were always better (six); and (5) that the program's funding specifically precluded non-U.S. citizens (five).

Information was also obtained about the countries of origin of the fellows. Table 1 shows this information by country of origin and region of the world (although British Commonwealth countries other than Canada are classified together). The 22 fellows identified in this study were from all over the world apart from the African and Eastern-bloc countries.

Table 1

Origins of the 22 Former or Current Foreign Family Practice Fellows in the United States, 1988*

Origin	No. of Fellows	% of Fellows
Country		
Israel	5	23
Chile	3	14
Canada	2	9
United Kingdom	2	9
South Korea	2	9
Venezuela	2	9
Costa Rica	1	4.5
Japan	1	4.5
Columbia	1	4.5
Taiwan	1	4.5
New Zealand	1	4.5
Laos	1	4.5
Total	22	100
Area of world		
Latin America	7	31.8
Middle East	5	22.7
Far East	5	22.7
British Commonwealth†	3	13.7
Canada	2	9.1
Total	22	100

*In 1988 the authors surveyed the 49 U.S. fellowship programs in family medicine to investigate the past experience and future intentions of such programs toward non-immigrant foreign family physicians coming to the United States for fellowship training, and to obtain information about the origins of the fellows who had been, or were being, trained at that time. A total of 47 of the programs responded.

†British Commonwealth countries, except for Canada, are classified together.

Discussion

Although 21 of the 46 fellowship programs that responded to the second questionnaire stated that they would be willing to accept foreign trainees, six of those 21 were uncertain of their funding for foreign fellows. Only six had prior experience and a willingness to accept future trainees.

It is, however, encouraging that almost all of the programs with prior experience remained positive about accepting future fellows from other countries if funds were available.

No information was available on how many foreign family physicians apply for fellowships in the United States or what the perceived need for this training is in various countries.

Training teachers of family medi-

cine has been a major goal of family practice organizations in this country, with the intention of improving the quality and quantity of practicing family physicians. This need is as great in other countries, but their resources for training are much more limited.

This study identified two major understandable explanations for not accepting foreign fellows: a program having funding that specifically precludes non-U.S. citizens and a lack of adequate secure funds. Other reasons are less understandable. Language and licensing concerns may be referred to the Educational Commission for Foreign Medical Graduates (ECFMG), which has been specifically charged with the task of dealing with them. That U.S. graduates are always of higher quality is clearly an unjustified bias and misses the point of accepting foreign fellows. Also, the unique contribution the visiting fellow brings to the host program in representing different ideas and perspectives can be considerable.

The International Medical Scholars Program, co-ordinated by the ECFMG and sponsored by many of the most prestigious U.S. medical organizations, including the American Medical Association and the American Hospital Association, has recently been established. Its functions are expected to be those of reviewing and approving applications, facilitating placements, and providing stipend support where needed, after fellowship funds become available. This program should expedite the training of foreign family practice faculty but will be able to do so only if there are programs in the United States that are willing to accept these physicians. Our data indicate that at present there is a serious shortage of such programs.

References

1. American Academy of Family Physicians. *1985 AAFP Transactions*. Kansas City, Missouri: AAFP, 1985, p. 188.
2. Society of Teachers of Family Medicine. *Fellowship Directory*. Kansas City, Missouri: STFM, 1988.